ARMY RESEARCH LABORATORY

ARL-TR-4833
AD_____

Discrete Optimization of Electronic Hyperpolarizabilities in a Chemical Subspace

by B. Christopher Rinderspacher, Jan Andzelm, Adam Rawlett, Joseph Dougherty, David N. Beratan, and Weitao Yang

ARL-TR-4833 May 2009

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory
Aberdeen Proving Ground, MD 21005

ARL-TR-4833 **May 2009**

Discrete Optimization of Electronic Hyperpolarizabilities in a Chemical Subspace

B. Christopher Rinderspacher, David N. Beratan, and Weitao Yang
Duke University, Department of Chemistry

Jan Andzelm, Adam Rawlett, Joseph Dougherty,
Weapons and Material Research Directorate, ARL

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
May 2009	Progress	

4. TITLE AND SUBTITLE
Discrete Optimization of Electronic Hyperpolarizabilities in a Chemical Subspace

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
B. Christopher Rinderspacher, Jan Andzelm, Adam Rawlett, Joseph Dougherty, David N. Beratan, and Weitao Yang

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
ATTN: AMSRD-ARL-WM-MA
Aberdeen Proving Ground, MD 21005

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-4833

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

We introduce a general optimization algorithm based on an interpolation of property values on a hypercube. The resultant algorithm is related to branch-and-bound/tree-search methods. We apply the algorithm to the optimization of the first electronic hyperpolarizability for several tolane libraries. The search includes structural and conformational information. Geometries were optimized using the Austin Model 1 (AM1), and first hyperpolarizabilities were computed using Intermediate Neglect of Differential Overlap/Screened Approximation (INDO/S). Even for small libraries, a significant improvement of the hyperpolarizability, up to a factor of ca. 4, was achieved. The algorithm was validated for efficiency and reproduced known experimental results. The algorithm converges to a good, local optimum as the logarithm of the library size, making large libraries accessible. For larger libraries, the improvement was accomplished by performing electronic structure calculations on less than 0.01% of the compounds in the larger libraries. Alternation of electron donating and accepting groups in the tolane scaffold was found to produce the best candidates consistently.

15. SUBJECT TERMS
Hyperpolarizabilities, inverse design, ab initio, optimization, tolane

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			B. Christopher Rinderspacher
Unclassified	Unclassified	Unclassified	UU	30	19b. TELEPHONE NUMBER *(Include area code)* (410) 306-2238

Contents

List of Figures

List of Tables

Acknowledgments

We would like to thank S. Keinan, B. Desinghu, G. Lindsay, A. Chafin, and M. Davis for helpful discussions. We are thankful to the Defense Advanced Research Projects Agency (DARPA) Predicting Real Optimized Materials through the Army Research Office (ARO) (W911NF-04-1-0243) and the U.S. Army Research Laboratory (ARL) for funding.

1. Introduction

In recent years, organic molecules have garnered increasing attention as components of high-hyperpolarizability materials, partly due to the variety of synthetically accessible compounds (*1, 2*). Applications for materials with high hyperpolarizabilities are found in telecommunication (*3*). The nonlinear response of organic molecules often finds its origin in the conjugated π-system, which facilitates the electronic polarizability. The design of such molecules *in silico* is complicated by the fact that chemical space, even constrained to smaller organic compounds, is combinatorially complex. The number of organic molecules of medium size is estimated to be on the order of 10^{200} (*4*). Enumeration is therefore unfeasibly costly and other methods for property optimization need to be developed. Including conformational searching further complicates molecular design.

Methods for optimization in discrete spaces have been studied extensively and recently reviewed (*5*). Optimization methods include integer programming, as in branch-and-bound techniques (including dead-end elimination [*6*]), simulated annealing (*7*), and genetic algorithms (*8*). These algorithms have found renewed interest and application in molecular and materials design (*9–12*). Recently, new approaches have been explored to embed discrete chemical space in continuous spaces to take advantage of continuous optimization techniques. These include, in particular, activities in our group on the linear combination of atomic potentials (LCAP) (*13–15*) method and the approach of von Lilienfeld (*16–18*), using a grand-canonical ensemble strategy. Here, we further employ continuous optimization methods aimed at discovering structures with optimal properties.

The problem of discrete optimization in chemical space can be tackled by embedding the discrete space in a virtual continuous space, parameterized by a set of continuous variables. This strategy establishes a continuous path from one molecule to another. Such a space can be constructed by defining molecules as a succession of replacements of an atom or molecular fragment by another. These fragment or atom placements need only satisfy the rules of valency. For example, a hydrogen in CH_4 might be replaced by a halogen or a methyl group, each corresponding to a specific geometry (or ensemble of geometries), energy(ies), and property value(s). It is possible to construct a continuous transition between Hamiltonians for the chemical structures as was done for LCAP (*13*). Equation 1 illustrates the procedure.

$$H(\boldsymbol{\lambda}) = \sum \lambda_i H_i, \sum_i \lambda_i = 1, 0 \leq \lambda_i \leq 1 \forall i \tag{1}$$

Each Hamiltonian H_i acts only on its own molecular subspace Ω_i projecting all other wave-functions out, and H acts on the direct sum of these spaces $\bigoplus_i \Omega_i$. In equation 1, the summation constraint implies the mutual exclusivity of the groups in the library (e.g., in

the previous example, as the hydrogen component increases toward 1, the halogen component decreases toward 0). In this approach, the groups are still linked through the wave function. Therefore, it is possible that all optima are at non-physical configurations (e.g., half hydrogen and half halogen in the same location). Starting with each $\lambda_i \in \{0, 1\}$, it is possible to compute the numerical derivative of a property P. We now explore the application of this idea for discrete optimization of the first hyperpolarizability.

2. Methods

2.1 Linear Interpolation of Discrete Spaces

Analogous to LCAP optimization, any property can be interpolated in a virtual continuous space. We call the interpolated space "virtual" since non-integer λ_i-values correspond to intermediate or "alchemical" species. In general, given a library with N molecules with property values P_s for molecule s, $\log_2 N$ variables may be used to embed the discrete library in the continuous space $[0, 1]^{\log_2 N}$. For example, assume a library consisting of methane, ethane, propane, and butane in exactly that order (figure 1). It is possible to interpolate among the four molecules using the parameters λ_0 and λ_1. A (quadratic) polynomial interpolating the ground state energies (for example) is

$$\begin{aligned} E(\lambda_0, \lambda_1) =& E_0(1-\lambda_0)(1-\lambda_1) + E_1\lambda_0(1-\lambda_1) + \\ & E_2(1-\lambda_0)\lambda_1 + E_3\lambda_0\lambda_1 \end{aligned} \tag{2}$$

This energy equation has a well-defined minimum. Interpolation using a single variable for this set of compounds would produce a third degree polynomial, but homogeneous solutions to third order polynomials are not trivial, and the optimum is not guaranteed to correspond to a molecule, i.e., $\lambda \in \{0, 1, 2, 3\}$.

Molecule	s	$\lambda_1\lambda_0$
CH_4	0	0,0
C_2H_6	1	0,1
C_3H_8	2	1,0
C_4H_{10}	3	1,1

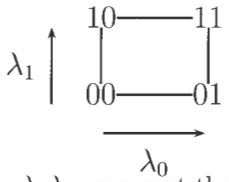


Figure 1. Simple example for interpolation. The bits $\lambda_1\lambda_0$ represent the molecule number $s = 2\lambda_1 + \lambda_0$ in the binary system.

The preceding example highlights the dependence of the property polynomial on the ordering of the molecules. Generalization of the example to a library $\mathcal{L}$ of size N leads to equations 3 and 4. Equation 3 describes the bit-string (binary) representation of a number s with bit $s(i)$ at the ith position:

$$s = \sum_i s(i) \cdot 2^i, s(i) \in \{0, 1\} \tag{3}$$

$$\tilde{P}(\boldsymbol{\lambda}) = \sum_{s=0}^{N-1} P_s \prod_{b=1}^{\log_2 N} ((1-\lambda_b)^{s(b)} \lambda_b^{1-s(b)}) \tag{4}$$

Equation 4 defines the property interpolation $\tilde{P}$ based on the bit-strings. We differentiate between $\tilde{P}$ and P to emphasize the domain of definition. The former is defined on the "virtual" space $[0,1]^{\log_2 N}$, while the latter is defined on the discrete space $\mathcal{L}$. This polynomial of the same order as it has variables ($\log_2 N$) is continuous on $[0,1]^{\log_2 N}$.

2.2 Derivatives of $\tilde{P}$

In order to use conventional optimization algorithms on continuous spaces, it is necessary to find the derivatives of $\tilde{P}$.

$$\frac{\partial \tilde{P}}{\partial \lambda_j}(\boldsymbol{\lambda}) = \sum_{s=0}^{N-1} P_s (-1)^{s(j)} \prod_{b \neq j}^{\log_2 N} ((1-\lambda_b)^{s(b)} \lambda_b^{1-s(b)}) \tag{5}$$

$$\frac{\partial^2 \tilde{P}}{\partial \lambda_k \partial \lambda_l}(\boldsymbol{\lambda}) = \sum_{s=0}^{N-1} P_s (-1)^{s(k)+s(l)} \prod_{b \notin \{k,l\}}^{\log_2 N} ((1-\lambda_b)^{s(b)} \lambda_b^{1-s(b)}) \tag{6}$$

Equations 5 and 6 show first and second order analytical derivatives of $\tilde{P}$. The derivative of $\tilde{P}$ at $\boldsymbol{\lambda}$ corresponding to the molecule with number s in the library $\mathcal{L}$ can be computed from nearest bit-string neighbors ($s^{(j)}$, $s^{(k,l)}$):

$$s^{(j)} = s + (-1)^{s(j)} \cdot 2^{s(j)} \tag{7}$$

$$s^{(k,l)} = s + (-1)^{s(k)} \cdot 2^{s(k)} + (-1)^{s(l)} \cdot 2^{s(l)} \tag{8}$$

$$\lambda_i = s(i), \frac{\partial \tilde{P}}{\partial \lambda_j}(\boldsymbol{\lambda}) = (-1)^{s(j)} (P_s - P_{s^{(j)}}) \tag{9}$$

$$\frac{\partial^2 \tilde{P}}{\partial \lambda_k \partial \lambda_l}(\boldsymbol{\lambda}) = (-1)^{s(k)} (-1)^{s(l)} (P_s - P_{s^{(k)}} - P_{s^{(l)}} + P_{s^{(k,l)}}), l \neq k, \lambda_i = s(i) \tag{10}$$

The highly nonlinear, but continuous description $\tilde{P}$ allows the development of optimization methods by substituting derivatives by finite differences in continuous optimization methods. In this case, the analytical property derivatives for a molecule are computed from simple (finite) property value differences, unlike in LCAP. The derivatives of LCAP need not be on straight lines pointing from one physical (non-"alchemical") molecule to another, although the property values of each real molecule are the same for either optimization scheme.

2.3 Comparison with Dead-End Elimination

To compare our approach (equation 4) with dead-end-elimination algorithms (DEE), we consider the minimization of a pairwise additive property function comprised of unary contributions $P_i^{(\mu)}$ acting on site i with occupation μ and binary contributions $P_{ij}^{(\mu,\nu)}$ acting on sites i and j with occupation μ and ν (equation 11):

$$P_s = \sum_i P_i^{(s(i))} + \sum_{i<j} P_{ij}^{(s(i),s(j))} \tag{11}$$

Collecting all terms, we find a quadratic dependence of $\tilde{P}$ on the pairwise terms P_{ij} with the parameters λ_i. Consequently, the derivatives are linear with respect to λ_i (equation 13):

$$\begin{aligned}\tilde{P}(\boldsymbol{\lambda}) = &\sum_i \left(P_i^{(0)}\lambda_i + P_i^{(1)}(1-\lambda_i)\right) + \\ &\sum_{i<j} \Big(P_{ij}^{(0,0)}\lambda_i\lambda_j + P_{ij}^{(1,0)}(1-\lambda_i)\lambda_j + \\ &\quad P_{ij}^{(0,1)}\lambda_i(1-\lambda_j) + P_{ij}^{(1,1)}(1-\lambda_i)(1-\lambda_j)\Big)\end{aligned} \tag{12}$$

$$\begin{aligned}\frac{\partial\tilde{P}}{\partial\lambda_i} = P_i^{(0)} - P_i^{(1)} + \sum_{j\neq i} \Big([P_{ij}^{(0,0)} - P_{ij}^{(1,0)}]\lambda_j + \\ [P_{ij}^{(0,1)} - P_{ij}^{(1,1)}](1-\lambda_j)\Big)\end{aligned} \tag{13}$$

From equation 13, a pruning argument for minimization, reminiscent of DEE, can be derived. Whenever the gradient with respect to a parameter λ_i is negative for all configurations of $\boldsymbol{\lambda} \in [0,1]^{\log_2 N}$ (equation 14), then $\lambda_i = 1$ minimizes $\tilde{P}$. This condition is only met when inequality 15 is fulfilled:

$$s(i) = 1 \Leftarrow \frac{\partial\tilde{P}}{\partial\lambda_i} < 0 \;\forall\lambda_j \in [0,1] \Leftrightarrow \tag{14}$$

$$P_i^{(0)} - P_i^{(1)} < \sum_{j\neq i} \min\{P_{ij}^{(1,0)} - P_{ij}^{(0,0)}, P_{ij}^{(1,1)} - P_{ij}^{(0,1)}\} \tag{15}$$

Conversely, a positive gradient implies that $\lambda_i = 0$ (equation 16) and the corresponding necessary and sufficient condition can be found in equation 17. Thus it has been demonstrated that $\tilde{P}$ naturally leads to DEE-like algorithms.

$$s(i) = 0 \Leftarrow \frac{\partial\tilde{P}}{\partial\lambda_i} > 0 \;\forall\lambda_j \in [0,1] \Leftrightarrow \tag{16}$$

$$P_i^{(0)} - P_i^{(1)} > \sum_{j\neq i} \max\{P_{ij}^{(1,0)} - P_{ij}^{(0,0)}, P_{ij}^{(1,1)} - P_{ij}^{(0,1)}\} \tag{17}$$

2.4 Library Construction and Ordering

The choice of enumeration of the library $\mathcal{L}$ determines the assignment of specific molecules to $\boldsymbol{\lambda}$. Consequently, this choice greatly influences the characteristics of $\tilde{P}$, such as its smoothness. Just exchanging the position of two neighboring molecules in the library changes the sign of the derivative at the corresponding $\boldsymbol{\lambda}$. If the Hessian of the pairwise-additive property function is positive-semi-definite, the corresponding $\tilde{P}$ is convex and optimization quickly reaches the global minimum. Using steepest gradient or Newton-Raphson algorithms locates property extrema (minima). It is beneficial to find an ordering of the library that produces a convex property surface. The linearity in each parameter λ_i implies convexity of $\tilde{P}$ with respect to that parameter.

Assuming that molecules of similar structure have similar properties, a measure of similarity may be used to decrease the ruggedness/convexity of $\tilde{P}$. One choice to facilitate smooth property surfaces is the enumeration of molecules by subsequent substitutions from a starting compound (figure 2). The substitutions may be defined recursively. Each level of a hierarchy of substitutions consists of a molecular fragment or atom to be connected to the next higher level, a list of substitution sites and a set of subsequent levels for each site (figure 2). Each element of the set of subsequent levels is identified with a coefficient between 0 and 1, and the sum of these coefficients for each set must equal 1 (see equation 18). For a case in which more than two possible substitutions are available at a site, the bit-string representation must be extended to allow mixed numeric bases b_k. The properties discussed above remain unchanged in this alternative interpolation (equation 20).

$$\sum_j \lambda_{ij} = 1, j \in \{0, \ldots, b_i - 1\} \tag{18}$$

$$s = \sum_i \left(\prod_{k=0}^{i-1} b_k \right) \sum_{j \in b_i} s(i,j) \cdot j, s(i,j) \in \{0,1\}, \sum_j s(i,j) = 1 \tag{19}$$

$$\tilde{P}\left(\{\lambda_{ij}\}_{j \in \{0,\ldots,b_i\},i}\right) = \sum_{s=0}^{N-1} P_s \left(\prod_i \prod_{j=0}^{b_i - 1} \lambda_{ij}^{s(i,j)} \right) \tag{20}$$

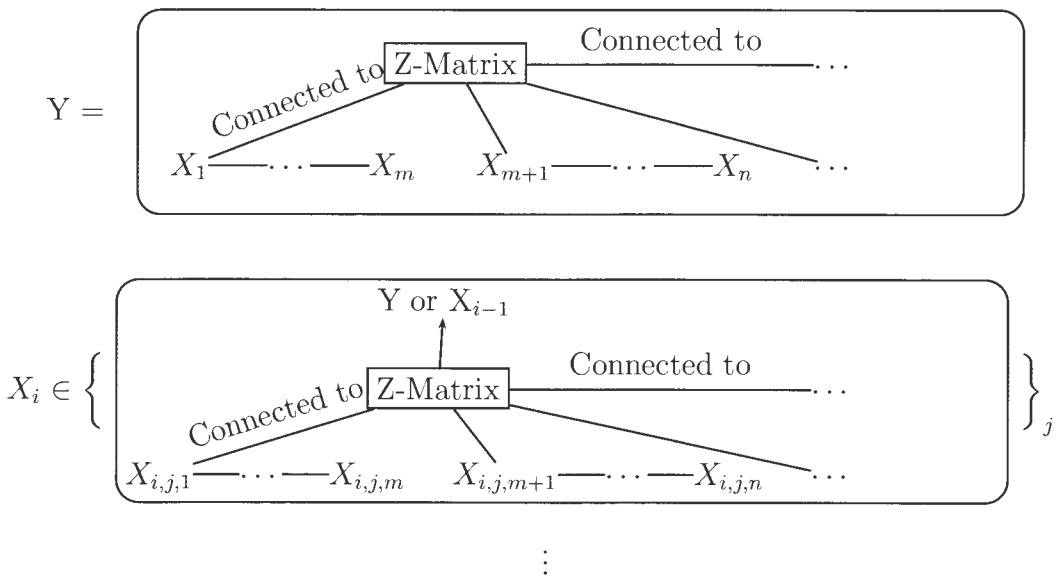


Figure 2. Substitution pattern hierarchy. Y contains a Z-matrix that has several open valences. The first can be filled with substituents found in X_1, which are connected substituents found in X_2, etc. The second is filled from X_m in the same manner. The X_i themselves are taken from a set of substitution patterns of the same kind as Y. Each instance is anchored to Y at the appropriate valence. The substitutions are terminated by Z-matrices that have no open valences.

2.5 Inclusion of Conformational Complexity

For each molecule, it is important to find low-energy conformers for the optimization to be physically meaningful. For each molecule in the molecular library, another optimization can be started with the (second) library consisting of the corresponding conformers. Each dihedral degree of freedom can be treated as a substitution site at the lowest level with a number of rotations as possible substitutions, as is commonly done in conformational searches (*6,19*). In this manner, the conformational search can be introduced as the lowest level in the previously described substitution hierarchy. Thus, the conformational search precedes property computation in property optimizations. More general constraints on the optimal molecule can be introduced via alternate methods, like Lagrange multipliers or stochastic algorithms. Lagrange multipliers can be implemented using (soft) penalty functions with weightings that increase throughout the optimization.

2.6 Algorithm

Here, a line search algorithm is used, in particular, each parameter λ_i is followed to a minimum in that direction before varying the next parameter λ_{i+1}. Maximization via this algorithm can be achieved, for instance, by minimizing the negative objective function. This line search algorithm is an implicit branch-and-bound algorithm. A flowchart for the

employed recursive algorithm appears in figure 3 and application of the algorithm to a small example will be discussed in section 3 under framework A (see also the accompanying figure 6).

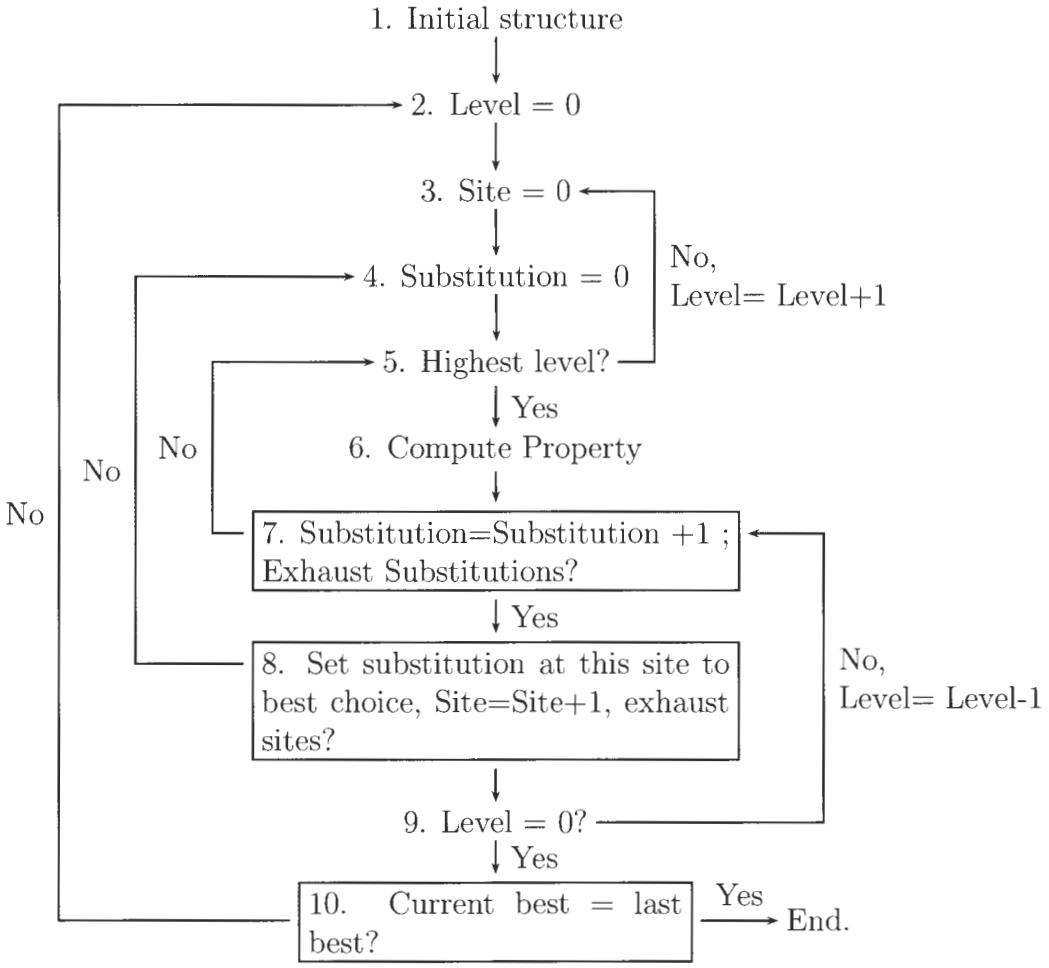


Figure 3. Flowchart of the algorithm.

Since $\tilde{P}(\boldsymbol{\lambda})$ is locally convex, this algorithm converges locally. The line-search steps 4–7 in figure 3 correspond to a linear tree search or branch-and-bound algorithm. The computational complexity is on the order $O(\log N)$ in the library size N due to the linear dependence on the $\log N$ variables. In contrast to conventional branch-and-bound methods, no structures are explicitly excluded from the search space. Since each molecule chosen in step 8 in figure 3 is strictly better in the sense of property optimization than its predecessor, the algorithm quickly converges to a local property value minimum in the library (*20*).

All property minima for this algorithm are minima for the steepest-descent derived method and vice versa. This algorithm traverses the library in a smoother fashion compared to the steepest-descent derived method, successfully employed by Keinan et al. (*15*), because the molecules are traversed variationally by single substitutions. While on one hand the steepest-descent based approach can sidestep barriers in the immediate vicinity efficiently,

due to the simultaneous change of potentially several bits, the variational nature of this line search guarantees convergence, which is particularly useful on rugged property surfaces.

For the sake of computational accessibility, all geometries were optimized using the semi-empirical Austin Model 1 (AM1) method as implemented in Gaussian03 (*21*). The static electronic hyperpolarizability was computed using Intermediate Neglect of Differential Overlap/Screened Approximation (INDO/S) as implemented in Complete Neglect of Differential Overlap (CNDO) by Reimers et al. (*22*) using the sum-over-states expression in equation 21. The configuration interaction (CI) space was spanned by up to 100 unoccupied or occupied orbitals to accommodate for the large number of electrons in some of the investigated systems.

$$\beta_{ijk} = \sum_{\nu\kappa} \frac{\langle 0| x_i |\nu\rangle \langle \nu| x_j - \mu_j |\kappa\rangle \langle \kappa| x_k |0\rangle}{E_{0\nu} E_{0\kappa}} \tag{21}$$

$$\beta_i = \frac{1}{3} \sum_j (\beta_{ijj} + \beta_{jij} + \beta_{jji}) \tag{22}$$

$$\beta_\mu = \frac{\vec{\mu}}{\|\vec{\mu}\|} \cdot \vec{\beta},\ \beta_0 = \left\|\vec{\beta}\right\| \tag{23}$$

where $E_{0\nu}$ is the excitation energy from the ground state to the νth excited state, $\vec{\beta}$ is the static electronic hyperpolarizability with components β_i and corresponding hyperpolarizability tensor elements β_{ijk}, β_0 is the isotropic hyperpolarizability, β_μ is the hyperpolarizability component in direction of the ground state dipole moment, $\vec{x}$ is the dipole operator with components x_i, and $\vec{\mu}$ is the ground state dipole moment with components μ_i.

Figures 4 and 5 summarize the tolane-based system studies. Tolane spectroscopic properties are favorable for applications, so their first and second hyperpolarizabilities have been studied extensively (*23, 24*). In addition, these structures are readily modified (*25*) and present a large number of possible derivatives. Tolanes, therefore, present a particularly rich testbed for these optimization studies.

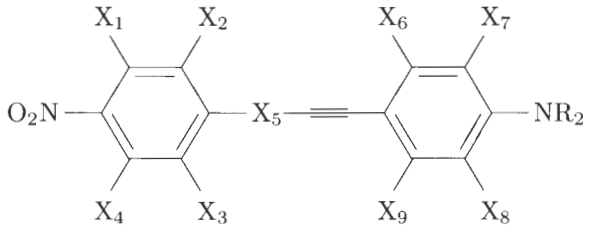
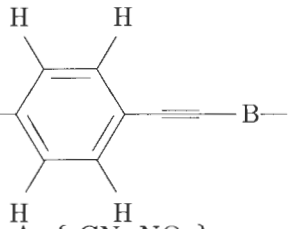
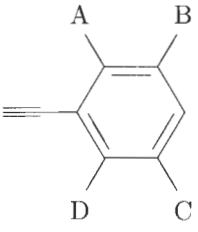


Figure 4. Tolane framework in which X_i and R_i are variable substituents.

A. Y= A—B′  B—D; B, B′ ∈ { o-, m-, p-, o′-, m′-phenyl}; D∈{OH, SH, NH_2, NMe_2 }; A∈{ CN, NO_2}

B. X_5 is empty; R=H; X_1,X_2,X_3,X_4,X_6,X_7,X_8,X_9 ∈ {CH_2OCH_3, CH_2OH, CH_2NH_2, NH_2, OH, NO_2}

C. X_5=  ;

(1) X_1=X_4=X_7=X_8=H; A, B, C, D, X_2, X_3, X_6, X_9 ∈ {CH_2OMe, CH_2OH, CH_2NH_2, NH_2, OH, CHO, NO_2}; R=CH_3

(2) X_1=X_4=X_7=X_8=H; A, B, C, D, X_2, X_3, X_6, X_9 ∈ {H, F, Cl, Br}; R=CH_3

(3) X_1=X_4=X_7=X_8=H; A, B, C, D, X_2, X_3, X_6, X_9 ∈ {CH_2OMe, CH_2OH, CH_2NH_2, NH_2, F, Cl, Br, CHO, NO_2}

Figure 5. Tolane libraries investigated.

3. Results and Discussion

Overall, five different tolane libraries were investigated (general structure in figure 4). The first three sets of molecules are optimized with respect to their static isotropic hyperpolarizability β_0 (equation 23), while the remaining sets are optimized with respect to the component of the hyperpolarizability in direction of the dipole β_μ (equation 23).

3.1 Framework A

Validation of the algorithm was performed on the structure framework A in figure 5. Figure 6 shows the progress of the algorithm. There are 200 molecules in this library, but hyperpolarizabilities of only 24 different molecules were computed during the optimization, the minimum number of molecules required for the algorithm to finish the optimization. Regardless of the starting structure, the algorithm consistently finishes with the global hyperpolarizability optimum (figure 6), which has also been confirmed experimentally (*26*). For comparison, if the library is searched randomly, the expected number of computed molecules before finding the global minimum is 200 molecules. If repeats are avoided, then still 101 molecules would need to be computed on average in order to obtain the same result.

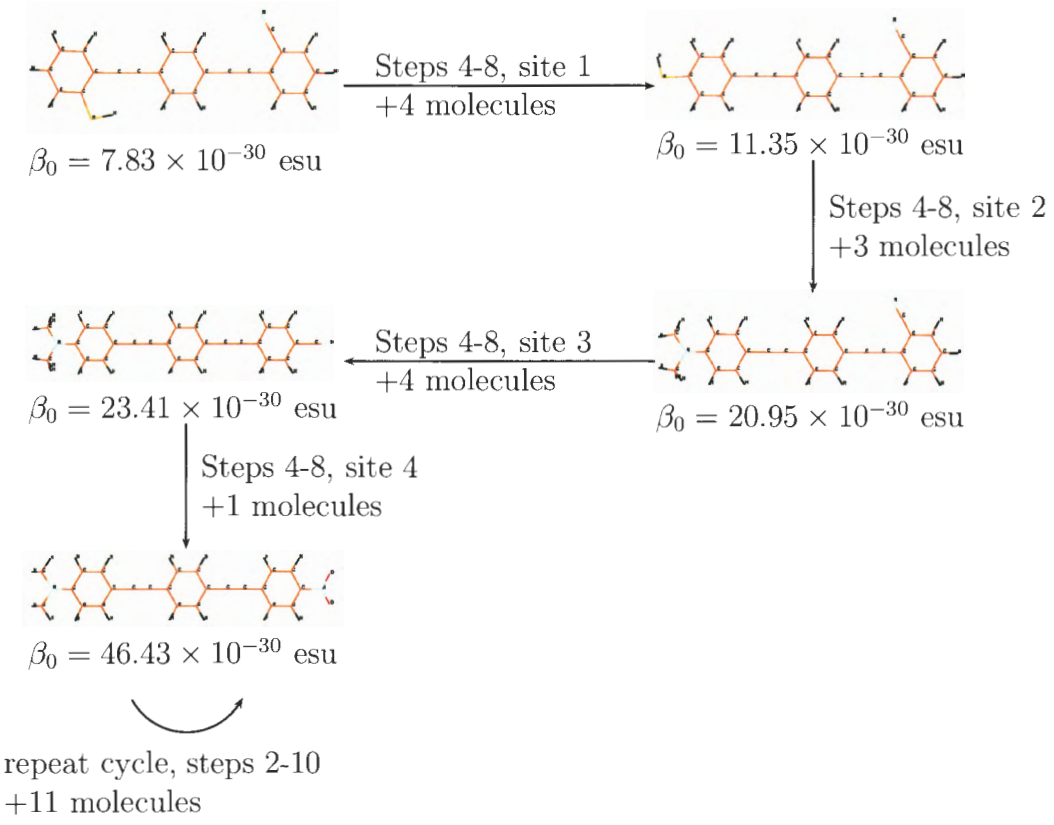


Figure 6. Progress of the optimization algorithm. The steps refer to the steps in figure 3. The number of molecules indicated is the number of previously unvisited molecules for which the property is computed in performing the steps. Carbons are marked in orange, hydrogens in white, oxygens in red, and nitrogens in light blue.

3.2 Framework B

The static hyperpolarizability β_0 of framework B in figure 5 optimizes to an unstable, perhaps explosive, structure with mostly nitro- and amino-substituents (figure 7). The final computed β_0-value was 131.9×10^{-30} esu after 121 computed structures from $6^8 \approx 1.7 \times 10^6$ possible molecules. Additionally, conformational analysis was performed. CHO and OH were allowed two possible orientations in the plane of the tolane. For CH_2OH and CH_2NH_2, three-fold rotation around the C-O and C-N bonds, respectively, was included, while only two-fold rotations around the bonds connecting to the tolane framework were allowed.

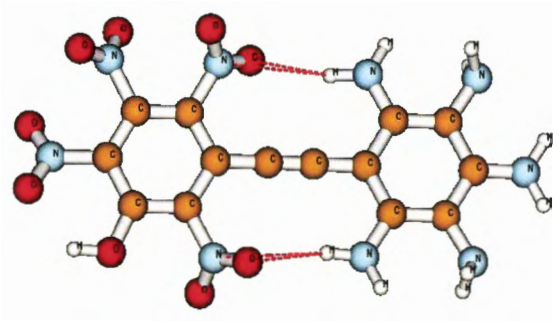


Figure 7. Final structure of framework B. Carbons are marked in orange, hydrogens in white, oxygens in red, and nitrogens in light blue.

3.3 Framework C-1

The static hyperpolarizability for compounds in C-1 of figure 5 was optimized starting from three different initial structures. A total of $7^8 \approx 5.8 \times 10^6$ possible molecules exist in this family. Conformational considerations were treated as in framework B. Two of the three runs converged to the same structure ($\beta_0 = 214.6 \times 10^{-30}$esu), while the third converged to a second structure with comparable hyperpolarizability ($\beta_0 = 216.9 \times 10^{-30}$esu, see tables 1 and 2). All three runs finished after computing less than 0.1% of all possible molecules and achieved three- to four-fold improvements of the hyperpolarizability. Comparing the two structures, some common motifs emerge: the variable fragments X_2 and X_3 contain nitro-groups, while X_6 and X_9 are occupied by amino-groups; furthermore, positions B and C are occupied by electron acceptors and sites A and D are occupied by electron donors. It is notable that not all positions are occupied by the "strongest" donors or acceptors in the substitution set, i.e., NH_2 and NO_2, respectively.

3.4 Framework C-2

Halogen substituents do not necessitate extensive conformational analysis, so they allow the evaluation of the optimization method without added constraints. The structures C-2 in figure 5 were optimized for the hyperpolarizability in the direction of the dipole moment (β_μ, see equation 23). Entries (a) and (c) in table 3 show the results of two optimizations of framework C-2 in figure 5 starting from the same initial structure with all substitutions set to hydrogens. In this case, convergence to a hyperpolarizability maximum is confirmed to be logarithmic in the library size, i.e., squaring the library size from 256 to 65536 leads to roughly twice the number of computed molecules.

Table 1. Starting and final structures of framework C-1 of figure 5. Carbons are marked in orange, hydrogens in white, oxygens in red, and nitrogens in light blue.

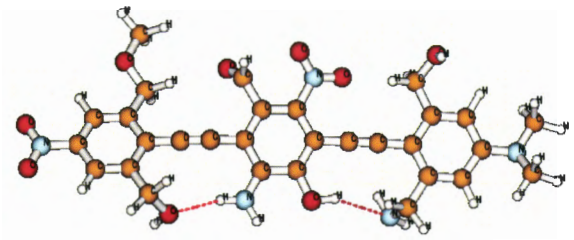
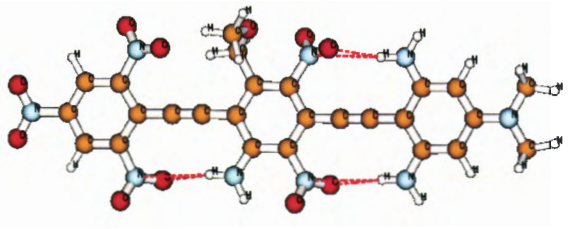
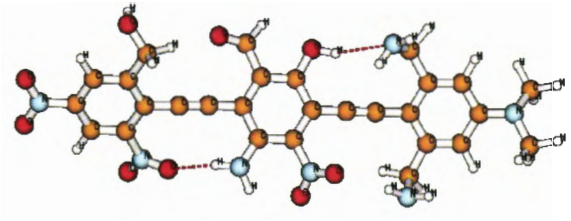
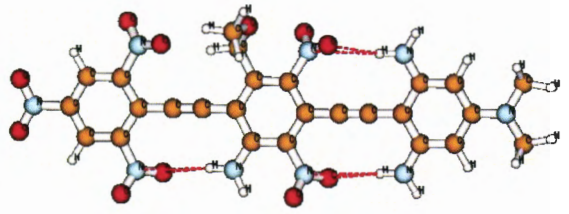
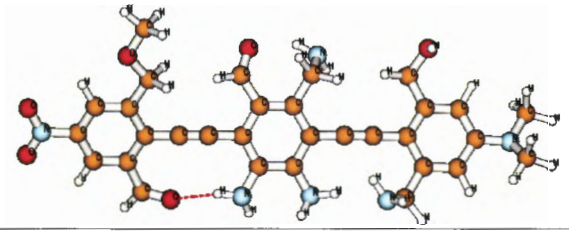
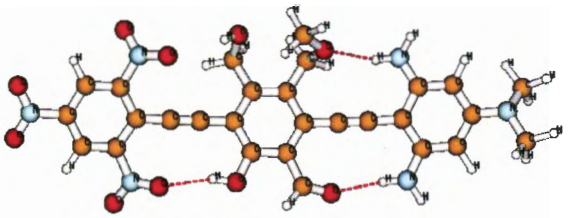
Run	Initial structure	Final structure
1		
2		
3		

Table 2. Starting and final hyperpolarizabilities and number of computed molecules for framework C-1 in figure 5.

Run	Initial $\beta_0/10^{-30}$esu	Final $\beta_0/10^{-30}$esu	Molecules Computed
1	55.1	214.6	157
2	71.0	214.6	109
3	49.9	216.9	169

Table 3. Optimized structures for frameworks C-2 in figure 5.

	Compound	$\beta_\mu/10^{-30}esu$	Molecules Computed	Library size
(a)	A,D,X_6,X_9=H; X_2,X_3=Br; B=Cl; C=F	84.1	67	65536
(b)	A,B,C,D,X_6,X_9=H; X_2,X_3=Br	77.4	69	65536
(c)	A,D=Br; X_6,X_9=H; X_2,X_3=Br; B,C=F	83.5	28	256
(d)	A,D=H; X_6,X_9=H; X_2,X_3=Br; B,C=Br	83.2	28	256

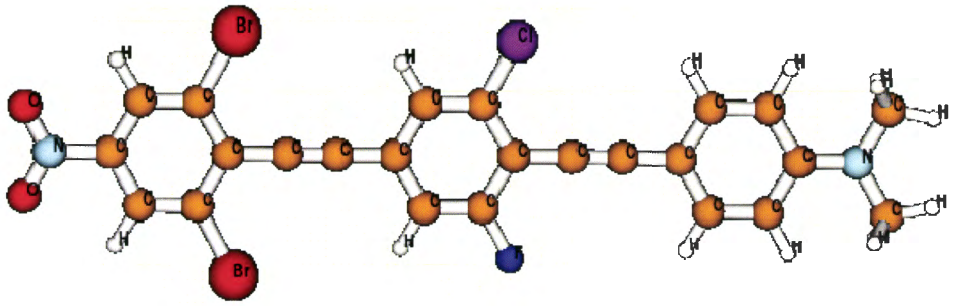


Figure 8. Largest β_μ structure for framework C-2 in figure 5. Carbons are marked in orange, hydrogens in white, oxygens in red, nitrogens in light blue, bromine in dark red, fluorine in dark blue, and chlorine in purple.

The stability of the optimization procedure was tested by constraining substitutions to be symmetric with respect to the mirror plane perpendicular to the plane of the backbone (runs (c) and (d) in table 3), as well as starting from different initial structures: runs (a) and (c) were started with all substituents set to hydrogen, while run (b) starts from X_2 =Br and X_8=F, and run (d) starts from $X_2 = X_3$ =Br and $X_7 = X_8$ =F. The hyperpolarizabilities of the initial structures were within 4 units of 50×10^{-30} esu. Since the procedure is not a global optimization algorithm, it is possible to end at different local maxima, here each run ended in a different structure with corresponding hyperpolarizabilities ($\beta_\mu/10^{-30}$ esu = 84.1, 77.4, 83.5, 83.2, respectively, see table 3). Nonetheless, the optimizations lead to significant and comparable improvements between runs. The found maxima all place bromine in the X_2 and X_3 positions, implying that a large fraction of the gain in β_μ arises from bromine to amino charge transfer interactions.

3.5 Framework C-3

Combining parts of libraries of C-1 and C-2 in figure 5, structures C-3 in figure 5 were subjected to optimization of the static hyperpolarizability in the direction of the dipole moment (β_μ). Four optimizations from different starting configurations were performed (see tables 4 and 5 for results). The "unbiased" first optimization leads to a five-fold increase in β_μ ($37.0 \rightarrow 181.5 \times 10^{-30}$esu). The final structure (see table 4) indeed is a mixture of the results for C-1 and C-2 in figure 5. The second optimization was started with a structure concentrating equal numbers of donors on one side and acceptors on the other, analogous to the final structure of framework B in figure 5. This starting structure exhibited only a marginally larger hyperpolarizability (55.6×10^{-30}esu) than the "unbiased" starting structure, but optimized to an alternating donor-acceptor arrangement (171.6×10^{-30}esu) that failed to reach the optimum found in the first optimization. The low hyperpolarizability is presumably due to the benzene rings twisting out of plane and reducing conjugation.

A biased starting point, with alternating donor and acceptor groups, leads to a marginally increased final hyperpolarizability (191.6×10^{-30}) over the first optimization. The attempt to exceed this value by substituting the "strongest" electron donors and acceptors, NH_2 and NO_2, fails despite the fact that this structure is indeed a local maximum (173.3×10^{-30}esu). All four optimization runs finish compute less than 0.001% out of the possible $9^8 \approx 4.3 \times 10^7$ molecules.

Table 4. Starting and final structures of framework C-3 in figure 5. Carbons are marked in orange, hydrogens in white, oxygens in red, nitrogens in light blue, bromine in dark red, fluorine in dark blue, and chlorine in purple.

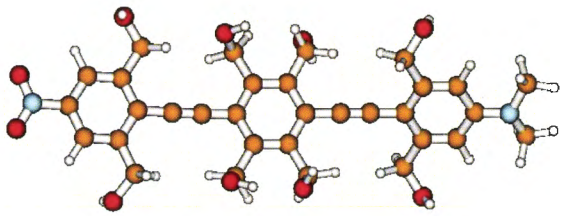
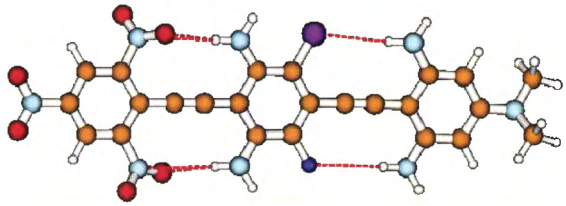
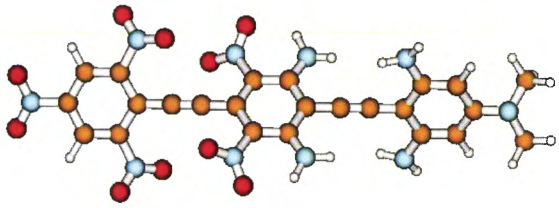
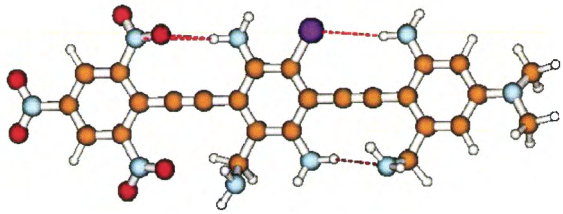
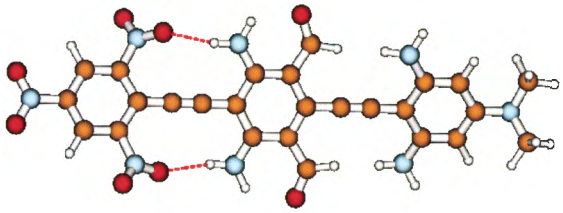
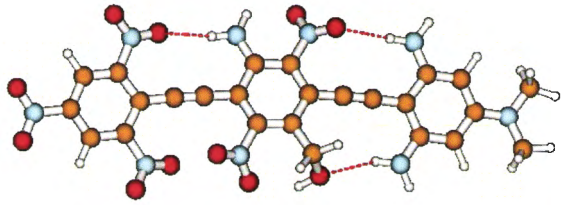
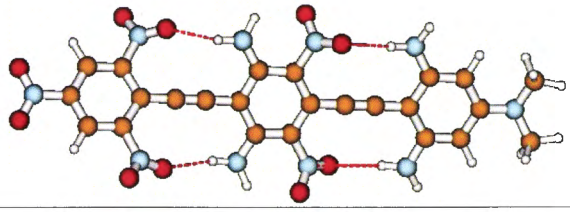
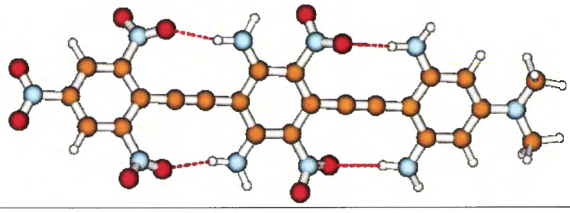
Run	Initial structure	Final structure
1		
2		
3		
4		

Table 5. Starting and final hyperpolarizabilities and number of computed molecules for framework C-3 in figure 5.

Run	Initial $\beta_\mu/10^{-30}$esu	Final $\beta_\mu/10^{-30}$esu	# Comp.
1	37.0	181.5	181
2	55.6	171.6	153
3	139.8	191.6	161
4	173.3	173.3	65

4. Summary and Conclusions

We have introduced an embedding of discrete molecular spaces in a continuous space, similar to the embedding of discrete Hamiltonians in LCAP (*27*). From this embedding, an optimization based on differentiation in the continuous space was developed. The theoretical framework transforms a discrete optimization problem into a continuous optimization problem, which then gives rise to a discrete optimization strategy. The theoretical complexity of the used line-search algorithm is $O(\log N)$ in the library size N and applications of the algorithm to a variety of conditions confirm the method's effectiveness. A design strategy for tolanes of alternating donors and acceptors along a conjugated framework is suggested by the optimization results. Further applications and improvements are under study including an extension to second-order derivative methods, probabilistic methods (*28*), and dynamic ordering of the parameters to achieve overall convexity.

References

[1] Andrekson, P. A.; Westlund, M. *Laser Photonics Rev.* **2007**, *1* (3), 231–248.

[2] Bergmann, G.; Ellis, C.; Hindmarsh, P.; Kelly, S. M.; O'Neill, M. *Mol. Cryst. Liq. Cryst.* **2001**, *368*, 4439–4446.

[3] Dalton, L. R.; Sullivan, P. A.; Bale, D. H.; Bricht, B. C. *3rd Nano and Giga Forum*, pp 1263–1277. Pergamon-Elsevier Science Ltd., 2007.

[4] van Deursen, R.; Reymond, J.-L. *Chem. Med. Chem.* **2007**, *2*, 636–640.

[5] Michalewicz, Z.; Fogel, D. B. *How to Solve It: Modern Heuristics.* Springer Verlag, Berlin, 2002.

[6] Gordon, D. B.; Mayo, S. L. *J. Comp. Chem.* **1998**, *19* (13), 1505–1514.

[7] Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220* (4598), 671–680.

[8] Muhlenbein, H.; Gorgesschleuter, M.; Kramer, O. *Parallel Comput.* **1988**, *7* (1), 65–85.

[9] Franceschetti, A.; Dudiy, S. V.; Barabash, S. V.; Zunger, A.; Xu, J.; van Schilfgaarde, M. *Phys. Rev. Lett.* **2006**, *97* (4).

[10] Dudiy, S. V.; Zunger, A. *Phys. Rev. Lett.* **2006**, *97* (4).

[11] Franceschetti, A.; Zunger, A.; van Schilfgaarde, M. *J. Phys.: Condens. Matter* **2007**, *19* (24).

[12] Piquini, P.; Graf, P. A.; Zunger, A. *Phys. Rev. Lett.* **2008**, *100* (18).

[13] Wang, M. L.; Hu, X. Q.; Beratan, D. N.; Yang, W. T. *J. Am. Chem. Soc.* **2006**, *128* (10), 3228–3232.

[14] Keinan, S.; Hu, X. Q.; Beratan, D. N.; Yang, W. T. *J. Phys. Chem. A* **2007**, *111* (1), 176–181.

[15] Keinan, S.; Paquette, W. D.; Skoko, J. J.; Beratan, D. N.; Yang, W. T.; Shinde, S.; Johnston, P. A.; Lazo, J. S.; Wipf, P. *Org. Biomol. Chem.* **2008**, *6* (18), 3256–3263.

[16] von Lilienfeld, O. A.; Lins, R. D.; Rothlisberger, U. *Phys. Rev. Lett.* **2005**, *95* (15).

[17] von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *J. Chem. Phys.* **2005**, *122* (1).

[18] von Lilienfeld, O. A.; Tuckerman, M. E. *J. Chem. Phys.* **2006**, *125* (15).

[19] Izgorodina, E. I.; Lin, C. Y.; Coote, M. L. *Phys. Chem.-Chem. Phys.* **2007**, *9* (20), 2507–2516.

[20] Desinghu, B.; Yang, W.; Beratan, D. N. *J. Chem. Phys.* **2008**, *129* (17), 174105.

[21] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Jr., J. A. M.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian 03; Technical report; Gaussian Inc.: Wallingford, CT, 2004.

[22] Tejerina, B.; Reimers, J. CNDO/INDO, 2008.

[23] Liu, C. P.; Liu, P.; Wu, K. C. *Acta Chim. Sinica* **2008**, *66* (7), 729–737.

[24] Oliva, M. M.; Casado, J.; Hennrich, G.; Navarrete, J.T.L. *J. Phys. Chem. B* **2006**, *110* (39), 19198–19206.

[25] Traber, B.; Oeser, T.; Gleiter, R. *Eur. J. Org. Chem.* **2005**, (7), 1283–1292.

[26] Nguyen, P.; Lesley, G.; Marder, T. B.; Ledoux, I.; Zyss, J. *Chem. Mater.* **1997**, *9*, 406–408.

[27] Xiao, D.; Yang, W.; Beratan, D. N. *J. Chem. Phys.* **2008**, *129* (4), 44106.

[28] Hu, X.; Beratan, D. N.; Yang, W. *J. Chem. Phys.* **2008**, *129* (6), 064102.

List of Symbols, Abbreviations, and Acronyms

AM1	Austin Model 1
ARL	U.S. Army Research Laboratory
ARO	Army Research Office
CI	configuration interaction
CNDO	Complete Neglect of Differential Overlap
DARPA	Defense Advanced Research Projects Agency
DEE	dead-end-elimination algorithms
INDO/S	Intermediate Neglect of Differential Overlap/Screened Approximation
LCAP	linear combination of atomic potentials

NO. OF COPIES	ORGANIZATION
1 ELEC	ADMNSTR DEFNS TECHL INFO CTR ATTN DTIC OCP 8725 JOHN J KINGMAN RD STE 0944 FT BELVOIR VA 22060-6218
1 HC	DARPA ATTN IXO S WELBY 3701 N FAIRFAX DR ARLINGTON VA 22203-1714
1 CD	OFC OF THE SECY OF DEFNS ATTN ODDRE (R&AT) THE PENTAGON WASHINGTON DC 20301-3080
1 HC	US ARMY RSRCH DEV AND ENGRG CMND ARMAMENT RSRCH DEV AND ENGRG CTR ARMAMENT ENGRG AND TECHNLGY CTR ATTN AMSRD AAR AEF T J MATTS BLDG 305 ABERDEEN PROVING GROUND MD 21005-5001
1 HC	PM TIMS, PROFILER (MMS-P) AN/TMQ-52 ATTN B GRIFFIES BUILDING 563 FT MONMOUTH NJ 07703
1 HC	US ARMY INFO SYS ENGRG CMND ATTN AMSEL IE TD F JENIA FT HUACHUCA AZ 85613-5300
1 HC	COMMANDER US ARMY RDECOM ATTN AMSRD AMR W C MCCORKLE 5400 FOWLER RD REDSTONE ARSENAL AL 35898-5000
1 HC	US GOVERNMENT PRINT OFF DEPOSITORY RECEIVING SECTION ATTN MAIL STOP IDAD J TATE 732 NORTH CAPITOL ST NW WASHINGTON DC 20402

NO. OF COPIES	ORGANIZATION
2 HCS	DUKE UNIVERSITY DEPT OF CHEMISTRY ATTN D BERATAN ATTN W YANG 124 SCIENCE DR BOX 90354 DURHAM NC 27708-0354
1 HC	US ARMY RSRCH LAB ATTN AMSRD ARL WM MA A RAWLETT BLDG 4600 RM C213 ABERDEEN PROVING GROUND MD 21005
1 HC	US ARMY RSRCH LAB ATTN AMSRD ARL WM MA C RINDERSPACHER BLDG 4600 RM C228 ABERDEEN PROVING GROUND MD 21005
1 HC	US ARMY RSRCH LAB ATTN AMSRD ARL WM MA J ANDZELM BLDG 4600 RM C204 ABERDEEN PROVING GROUND MD 21005
1 HC	US ARMY RSRCH LAB ATTN AMSRD ARL WM MA J DOUGHERTY BLDG 4600 RM C227 ABERDEEN PROVING GROUND MD 21005
1 HC	US ARMY RSRCH LAB ATTN AMSRD ARL CI OK TP T LANDFRIED BLDG 4600 ABERDEEN PROVING GROUND MD 21005-5066
1 HC	DIRECTOR US ARMY RSRCH LAB ATTN AMSRD ARL RO EV W D BACH PO BOX 12211 RESEARCH TRIANGLE PARK NC 27709

NO. OF COPIES	ORGANIZATION
3 HCS	US ARMY RSRCH LAB ATTN AMSRD ARL CI OK PE TECHL PUB ATTN AMSRD ARL CI OK TL TECHL LIB ATTN IMNE ALC HRR MAIL & RECORDS MGMT ADELPHI MD 20783-1197

TOTAL: 19 (1 ELEC, 1 CD, 17 HCS)

INTENTIONALLY LEFT BLANK.